Loss of *CFHR5* function reduces the risk for age-related macular degeneration

Mary Pat Reeve[1,2,3,*], Stephanie Loomis[4,*], Eija Nissilä[5], Tobias Rausch[6], Zhili Zheng[2,3], Pietro Della Briotta Parolo[2,3], Daniel Ben-Isvy[3,7,8], Elias Aho[5], Emilia Cesetti[5,9], Yoko Okunuki[4], Helen McLaughlin[4], Johanna Mäkelä[10], FinnGen, Mitja Kurki[1,2,3], Michael E. Talkowski[3,7], Jan O. Korbel[6], Kip Connor[4], Seppo Meri[5], Mark J. Daly[1,2,3,$,#], Heiko Runz[1,4,6,$,#]

1. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
2. Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
3. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA
4. Research and Development, Biogen Inc., Cambridge, MA, USA
5. Department of Bacteriology and Immunology, Translational Immunology Research Program, University of Helsinki, Helsinki, Finland
6. European Molecular Biological Laboratories (EMBL), Heidelberg, Germany
7. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
8. Division of Medical Sciences, Harvard Medical School, Boston, MA, USA
9. Department of Biomedical Sciences, Humanitas University, Milan, Italy
10. Finnish Biobank Cooperative (FinBB), Turku, Finland

* joint contribution
$ joint supervision

communication: heiko.runz@gmail.com or mjdaly@broadinstitute.org

Abstract

Age-related macular degeneration (AMD) is a prevalent cause of vision loss in the elderly with limited therapeutic options. A single chromosomal region around the complement factor H gene (*CFH*) is reported to explain nearly 25% of genetic AMD risk. Here, we used association testing, statistical finemapping and conditional analyses in 12,495 AMD cases and 461,686 controls to deconvolute four major *CFH* haplotypes that convey protection from AMD. We show that beyond *CFH*, two of these are explained by Finn-enriched frameshift and missense variants in the *CFH* modulator *CFHR5*. We demonstrate through a FinnGen sample recall study that *CFHR5* variant carriers exhibit dose-dependent reductions in serum levels of the *CFHR5* gene product FHR-5 and two functionally related proteins at the locus. Genetic reduction in FHR-5 correlates with higher preserved activities of the classical and alternative complement pathways. Our results propose therapeutic downregulation of FHR-5 as promising to prevent or treat AMD.

Introduction

Age-related macular degeneration (AMD) is characterized by a slowly progressive loss of central vision due to the death of retinal pigment epithelium and photoreceptor cells in the retina. The prevalence of AMD increases with age, and in Western countries it affects one in five individuals over the age of ninety[1]. While efficient treatments exist for the neovascular form of AMD (wet AMD), the medical need remains substantial for the much more common dry AMD, which distinguishes itself by macular atrophy and a progressive buildup of subretinal deposits known as drusen.

Twin and genome-wide association studies (GWAS) have established that up to 71% of AMD susceptibility is heritable[2,3], with the most recent meta-analyses identifying over 60 genome-wide significant loci[4,5]. Out of these, a single chromosomal region around the complement factor H gene (*CFH*) has been reported to explain up to 25% of genetic risk for dry and wet AMD[3]. This association, together with other loci including *CFI* and *C3*, have consolidated a central role of the complement system in AMD pathogenesis. By now, multiple lines of evidence propose that the overactivation of the alternative pathway C3 convertase, on which the classic, alternative and MBL-lectin complement pathways converge, initiates a proinflammatory cascade that results in retinal tissue damage. Activation of C3 can be inhibited by Factor H (FH), the protein product of *CFH* which slows down complement activation. Elevating FH has been suggested as a strategy for the treatment of AMD[6]; however, this approach is challenged by the need to keep regional FH levels within a narrow margin for an optimal efficacy and safety profile.

CFH was the first disease susceptibility locus ever unveiled through the GWAS approach[7-10]. Since its discovery, the complexity of this locus has increased with increasing GWAS sample sizes and geographic representation. Conclusive evidence of two very strong, independent variants at *CFH*[11,12] was published shortly after discovery, including the original association to the common *CFH* missense variant p.Tyr402His which in functional studies was shown to affect binding of FH to its interaction partners and increase AMD risk[13,14]. Similarly, rare loss-of-function variants in *CFH* such as p.Arg1210Cys were later demonstrated to accelerate the onset of AMD by several years[15]. However, *CFH* variants modulate susceptibility to AMD in conjunction with different haplotypes that include *CFH* together with a set of five adjacent and highly homologous genes termed *CFH* related genes 1-5 (*CFHR1-5*)[7,16,17]. The products of *CFHR1-5*, FHR1-5, are believed to compete with FH for C3b binding and counteract its attenuating effect on complement

activation[18,19]. Consistently, structural variants (SVs) resulting in deletions of *CFHR1* and *CFHR3*[16] or of *CFHR1* and *CFHR4*[20], respectively, have been proposed to reduce AMD risk, although the independence of such effects from *CFH* have remained contested[21,22]. Recently, through a coding variant meta-analysis in over 650,000 individuals we identified a frameshift variant within *CFHR5* (p.Glu163insAA, rs565457964) as associated with a strong protection from AMD[23], although also for this signal no independence from other variation at the locus has yet been established.

Here, we explored the relative contribution of *CFHR5* to the risk for AMD and its subforms through refined association testing and finemapping analyses for AMD in FinnGen (FG), a large population biobank cohort in Finland[24]. We show that genetic loss of *CFHR5* function is associated with a reduced risk for AMD irrespective of other variation at the *CFH* locus. We further demonstrate that *CFHR5* frameshift variant carriers not only exhibit reduced blood levels of FHR-5, but also of FHR-2 and FHR-4 proteins. Moreover, carriers have a higher capacity to activate the classical and alternative complement pathways, consistent with enhanced abilities of the complement system to clear retinal debris. Our results establish downregulation of *CFHR5* as an attractive opportunity for targeted AMD therapies.

Results

FinnGen GWAS reveals 23 regions independently associated with AMD.

We conducted an association study for AMD (https://risteys.finregistry.fi/endpoints/H7_AMD, including both wet and dry AMD) in FinnGen (FG) data freeze (DF) 12 (12,495 cases, 461,686 controls) across the full allele-frequency spectrum. This identified genome-wide significant associations to SNPs at 23 independent genomic regions (**Extended Data Figure 1**; **Supplementary Table 1**). 17 of these loci fall into regions that had been reported previously as associated with AMD or its subforms[3], while six appear to be novel. Systematic finemapping and conditional analyses prioritized likely causal variants for three out of five regions with multiple apparently independent signals. This included a single base pair insertion in an intron of *LIPC* (15:58430391:G:GC; p_{cond}=7.92E-08; minor allele frequency [MAF]=0.23) that nominates *LIPC* over *ALDH1A2* as the likely causal gene in an established AMD region on chromosome 15. Our analysis confirmed variants at the *HTRA1*/*ARMS2* locus (chr10:120954932-123954932) as the genome-wide strongest drivers of genetic association with AMD.

FinnGen confirms CFH regional variation as protective for AMD.

Consistent with previous AMD GWAS, we identified strong conditionally independent associations near several genes that encode members of the complement pathway, specifically *CFH*, *C3* and *CFI*. Out of these, *CFH* showed the second strongest association with AMD in Finns, with nearly 100 regional SNPs exhibiting p-values below 10^{-375}. Multiple studies have robustly established two independent variants in *CFH* as key contributors to this signal, one being tagged by rs1410996 ("$FG_{AMD}1$") and the other by the coding missense variant p.Tyr402His ("$FG_{AMD}2$")[11,12]. Consistently, proxies for both variants also showed the strongest signals in our analysis, with p-values of $9.6x10^{-618}$ and $2.0x10^{-590}$, respectively, in univariate analysis (**Figure 1**; **Supplementary Table 2**). We also replicated our earlier[23] of a strong association between the low-frequency frameshift variant p.Glu163insAA in *CFHR5* (chr1:196994128:C:CAA, rs565457964, $CFHR5_{fs}$) with protection from AMD ($p=1.1x10^{-68}$ in univariate analysis) at this locus. Notably, a phenome-

wide scan in carriers of this frameshift variant, which is ~4-fold enriched in Finns relative to UKB, in the latest FG release revealed significant associations only with protection from AMD and its subforms, but none of the 2,404 other phenotypes ascertained in FG DF12 (**Extended Data Figure 2**, **Supplementary Table 3**). Such lack of evident signs for adverse events in humans contributes to make therapeutic reduction of *CFHR5* or its products a potentially compelling strategy for the prevention or treatment of AMD, which motivated us to further de-risk this putative drug target through refined genetic and functional studies.

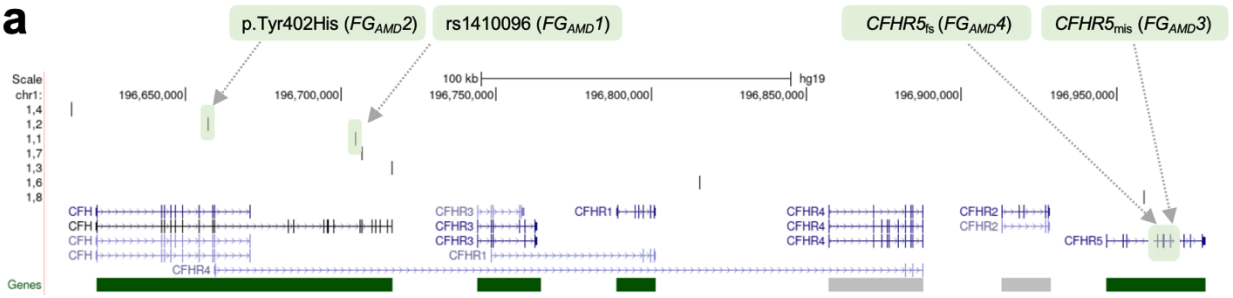
Finemapping CFH region reveals conditionally independent protective CFHR5 variants.

In order to determine whether $CFHR5_{fs}$ or any other regional variants were independently associated with AMD risk or protection, we performed conditional analyses in FG using REGENIE on a 1.15 Mb region bounded by recombination hotspots (chr1:196,200,000-197,350,000; hg38) and spanning 12 genes (*KCNT2*, *CFH*, *CFHR3*, *CFHR1*, *CFHR4*, *CFHR2*, *CFHR5*, *F13B*, *ASPM*, *SEPT14P12*, *ZBTB41* and *CRB1*). For consistency with an earlier AMD GWAS, we fixed the same top two variants identified in Fritsche et al., 2016[3], rs10922109 and rs570618, which are in complete LD (r2>0.99) with the originally reported variants rs1410996 and p.Tyr402His, respectively, to begin the conditional search for independently associated variants. Unsurprisingly, both SNPs explained the majority of AMD signal in the region also in our analysis. However, conditioning on both variants left a number of tertiary variants with conditional p-values between $5.9x10^{-7}$ and $6.2x10^{-25}$ (**Figure 1**; **Supplementary Table 2**) within a smaller ~380kb region encompassing *CFH* and all five *CFHR* genes. Within this region we found a third independent signal (p_cond=$6.2x10^{-25}$, beta=-0.50, "$FG_{AMD}3$") tagged by rs537634973, a roughly 2% allele frequency (AF) SNP enriched in Finns and in high LD with five other similarly associated variants, including missense variants in *CFHR5* (rs139017763, p.Gly278Ser, r^2=0.91) and *CFHR2* (rs79351096, p.Cys72Tyr, r^2=0.85). Notably, *CFHR5* p.Gly278Ser is also in high LD (r^2=0.72) with the fourth-strongest signal reported as independently driving the regional AMD association in Fritsche et al., 2016[3].

We repeated the conditional analysis including this third signal. This led to another genome-wide significant signal ($p_{cond}=6.8x10^{-11}$, beta=-0.29, "$FG_{AMD}4$") tagged by rs572896741 and in high LD with five other regional variants of ~4% AF and enriched up to 10-fold in FG relative to UKB. Notably, this fourth signal also included $CFHR5_{fs}$ (r^2=0.91) that, like *CFHR5* p.Gly278Ser, was associated with protection from AMD, indicating a potential allelic series in *CFHR5* that could be leveraged to develop targeted therapies.

No other signals within the 1.15Mb *CFH* region reached genome-wide significance after conditioning on all four signals, although we confirm several additional signals reported earlier[3] as conditionally independent beyond these four signals with $p<1x10^{-6}$ (**Supplementary Table 2**).

a



b

signal	GWAS lead	MAF case	MAF cont	FG lead	imp qual	univariate	p-value cond 1	p-value cond 2	conclusion	notes
1,1	rs10922109	0,223	0,426	rs10922109	>0.99	9.6×10^{-618}	...	...	**TRUE (FG1)**	r2 >.99 to *CFH* rs1410996
1,2	rs570618	0,58	0,364	rs570618	>0.99	2.0×10^{-590}	6.0×10^{-112}	...	**TRUE (FG2)**	r2 >.99 to *CFH* Y402H
1,3	rs121913059	0,003	0,00014	not imputed		8.9×10^{-24}	1.8×10^{-32}	n.i.	TRUE	*CFH* R1210C (rare)
1,4	rs148553336	0,003	0,009	rs537634973	0,98	8.6×10^{-26}	8.9×10^{-21}	6.2×10^{-25}	**TRUE (FG3)**	r2 =.91 to *CFHR5* missense
1,5	rs187328863	0,054	0,028	rs572896741	0,98			6.8×10^{-11}	**TRUE (FG4)**	r2 =.91 to *CFHR5* fs
				rs79436252	0,98	1.1×10^{-68}	9.2×10^{-13}	2.0×10^{-7}		nearest *CFHR5* fs proxy (n.i.)
1,6	rs61818925	0,284	0,385	rs61818925	0,98	6.0×10^{-165}	1.3×10^{-10}	0.29	p >.05	no independent signal
1,7	rs35292876	0,021	0,009	rs35292876	0,99	8.2×10^{-37}	2.9×10^{-9}	5.9×10^{-7}	p ~ 10^{-7}	proxy near signif., likely TRUE
1,8	rs191281603	0,007	0,006	rs191281603	0,94	0,68	1.4×10^{-8}	0,093	p >.05	no signal in FG univ or cond.

Figure 1. Conditional analysis of the *CFH* region reveals four major independent AMD signals. Finemapping and conditional analyses of genome-wide significant lead signals (FG lead) within the ~380kb *CFH* region (chr1:196,640,000-197,020,000; hg38) from a FinnGen (DF12) genome-wide association study for age-related macular degeneration (AMD) with 12,495 cases and 461,686 controls (see **Supplementary Table 1**). **(a)** Positions of coding genes and lead SNPs (1.1-1.8). Conditionally independent signals (FG_{AMD}1-4) are highlighted. **(b)** Association p-values from univariate and two rounds of conditional (cond.) analyses. "True" associations representing the major conditionally independent signals are highlighted in green. For details, see **Supplementary Table 2**. "GWAS lead" is based on Fritsche et al. 2016[3]; MAF, minor allele frequency; imp qual, imputation quality; n.i., not imputed.

No evidence for an impact of regional structural variants on AMD risk in Finns.

The *CFH* gene family resides within a genetically unstable region that has evolved from several incomplete segmental duplication events and harbors structural variants (SVs) of variable population frequency[20]. Several of these SVs have been reported to reduce or increase the copy numbers of one or more *CFHR* genes[8,20]. To ensure that SVs of potential relevance to the association evidence were adequately considered in our conditional analyses, we leveraged an SV map compiled from whole genome sequencing (WGS) data in more than 64,000 individuals from gnomAD (v4 release) and identified SNP proxies for two well-established *CFH* region deletion variants, a common deletion of 91kb encompassing *CFHR1* and *CFHR3*, and a low-frequency, Finn-enriched 121kb deletion of *CFHR1* and *CFHR4*. Importantly, neither of these two SVs were in LD with the credible sets used in finemapping, nor did they show a residual signal

after our stepwise conditional analysis. For instance, conditioning on $FG_{AMD}1$ reduced the strength of association between the FG lead variant rs61818925 (signal 1.6 in **Figure 1b**), a tight proxy for the *CFHR1*/*CFHR3* deletion (r^2=0.98), and AMD from $6.0x10^{-165}$ to $p=1.3x10^{-10}$, and further to p=0.29 upon jointly conditioning on $FG_{AMD}1$ and $FG_{AMD}2$. Our conditional analyses thus strongly propose that the often-studied large deletions in the *CFH* region do not substantively influence AMD risk in Finns. Because these are common deletion variants, this provides little support for the hypothesis that genetic variants impacting *CFHR1*, *CFHR3* and *CFHR4* contribute substantially to the strong association between the *CFH* region and AMD - which our results suggest are primarily explained by variants in *CFH* and *CFHR5* genes.

No evidence for cryptic regional variation to explain $CFHR5_{fs}$ effects.

To further validate whether *CFHR5* variants are indeed independently protective for AMD or potentially act in conjunction with yet concealed novel variation at the locus, we next leveraged phased short-read and long-read sequencing data from the expanded 1000 Genomes (1KG) project[25,26] and studied the *CFH* regional architecture at base-level resolution across ethnicities. We identified 37 individuals carrying a total of 38 *CFHR5* coding variants among the 6,404 chromosomes 1 haplotypes in this cohort (**Supplementary Table 4**). Among these, eight carried the p.Glu163insAA ($CFHR5_{fs}$) variant in a heterozygous and one in a homozygous state. Six were heterozygous carriers of another frameshift variant at the identical position, p.Glu163insA, and 22 carried the missense variant p.Gly278Ser. We further identified 12 carriers of a rare *CFHR5* intronic variant, rs191281603-G, that was proposed earlier as associated with AMD[17], but that did not meet independent significance in our conditional analyses. Consistent with our previous findings[23], *CFHR5* p.Glu163insAA was substantially more prevalent in unrelated Finns (7/198; AF=3.54%) relative to all other 1KG ethnicities (3/6,208; AF=0.048%; 73-fold enrichment), with the only non-Finnish carriers being two Tuscans from Italy and one African American individual. Also p.Gly278Ser showed a 12-fold enrichment in Finns (6/198; AF=3.03%) relative to the overall cohort (16/6,208; AF=0.26%), while only one of the rs191281603-G and none of the p.Glu163insA carriers were Finnish.

Interrogation of the *CFH* region (chr1:196,640,000-197,020,000; hg38) with phased WGS data across all available samples identified 10 SVs with AF>1% (**Figure 2a**; **Extended Data Figure 3**; **Supplementary Table 5**). This included three common SVs: the above-studied 91kb *CFHR1*/*CFHR3* deletion (HGSV_15069; AF=27.2%) as well as two insertions of 97bp (HGSV_15081; AF=20.7%) and 280bp (HGSV_15100; AF=20%), respectively. Rarer SVs in the region stretched from the above-studied *CFHR1*/*CFHR4* deletion (HGSV_15086; AF=1.9%) to a duplication of 64bp (HGSV_15114; AF=1.8%). None of these SVs impacted coding regions of *CFH* or *CFHR1-5* genes.

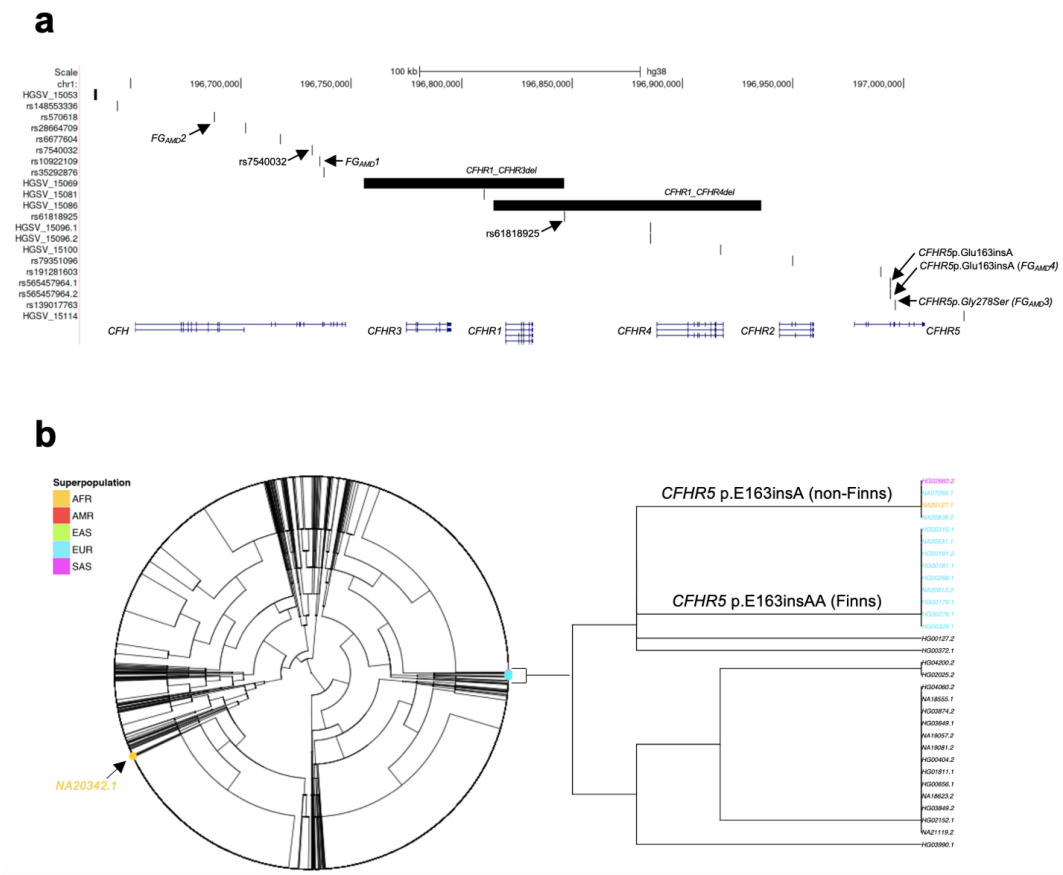


Figure 2. Phased short- and long-read sequencing excludes structural variants as a driver of *CFHR5* AMD associations. (a) Positions of coding genes, lead AMD GWAS SNPs (from this study and Lores-Motta et al. 2021[17]) and structural variants (SVs) with allele frequencies >1% identified through short- and long-read sequencing of the ~380kb *CFH* region (chr1:196,640,000-197,020,000; hg38) in 6,404 chromosomes of the 1,000 Genomes cohort (1KG). Previously described large deletions encompassing *CFHR1* and *CFHR3* (HGSV_15069), as well as *CFHR2* and *CFHR4* (HGSV_15086), respectively, are highlighted together with their proxy SNPs and identified coding variants in *CFHR5*. (b) Dendrogram of *CFH* region haplotypes in 1KG. Carriers of *CFHR5* frameshift variants and their ethnicity are highlighted. Haplotype 1 and 2 are indicated by sample suffixes, i.e., NA20342.1 corresponds to haplotype 1 of sample NA20342. For details, see **Supplementary Tables 4-8**.

Taking into account our FG lead signals, SVs with AF>1% and eight lead SNPs proposed as independent from published AMD GWAS[3,17], we identified 254 different haplotypes in the *CFH* region across the full 1KG cohort (**Figure 2b**; **Supplementary Tables 6-8**). This pleiotropy could be traced to just three foundational haplotypes that in Finns showed frequencies of 15%, 40% and 38%, respectively. Phased long-read data confirmed that the remaining 7% haplotypes are derivatives of the most common Finnish haplotype and contained either *CFHR5*$_{fs}$ or *CFHR5* p.Gly278Ser. Notably, despite being found exclusively in non-Finnish individuals, the haplotype structure of p.Glu163insA carriers was identical to that of *CFHR5*$_{fs}$ and *CFHR5* p.Gly278Ser carriers, suggesting that similar mutational events must have led to their occurrence. Nevertheless, also other evolutionary mechanisms may exist for *CFHR5* coding variants to arise, as evidenced by the single 1KG participant of African ancestry (NA20342) who carried *CFHR5*$_{fs}$ on an alternative haplotype.

Importantly, haplotype-resolved long-read sequencing in 1KG identified no additional cryptic rare or common variation that would be expected to modify protein-coding elements within the *CFH* region. Taken together, our base-level interrogation of the *CFH* region excludes that the reduced risk for AMD in *CFHR5* frameshift and missense variant carriers is conveyed by local SVs or concealed single base pair changes other than those impacting *CFHR5*.

CFHR5 frameshift variant carriers show lower FHR-5, FHR-2 and FHR-4 blood levels.

To test for a relevance of *CFHR5*$_{fs}$ on FHR-5 levels and function, we established a process to recall samples from broadly consented FG participants in a customized manner (**Extended Data Figure 4**). In partnership with the Finnish Biobank Cooperative (FinBB), we selected serum samples from 200 individuals with a diagnosis of dry AMD based on registry information that had been archived at four Finnish biobanks. We matched these cases to 200 controls based on age, sex, *CFH* and *CFHR5* variant status and the absence of alternative diagnoses affecting the eye (**Table 1**; **Supplementary Table 9**). The final recall study cohort had an average age of 73.6±9.1 and 74.4±8.5 years at diagnosis and blood sampling, respectively. Thirty-two participants (8%) were homozygous for the *CHFR5*$_{fs}$ variant, 85 (21%) were heterozygous and 282 (71%) were non-carriers. No AMD cases homozygous for *CHFR5*$_{fs}$ were available for resampling.

	Biobank				Total
	Auria (N=81)	Borealis (N=109)	Eastern Finland (N=101)	Tampere (N=108)	(N=399)
N AMD cases	39	60	49	52	199
Mean age at sampling (SD) [years]	70.4 (10.0)	76.6 (7.6)	74.0 (7.1)	75.5 (8.3)	74.4 (8.5)
Min; max	47; 91	61; 91	62; 92	56; 96	47; 96
Mean age at AMD diagnosis (SD) [years]	75.6 (10.5)	74.3 (8.3)	72.4 (7.1)	72.4 (10.2)	73.6 (9.1)
Min; max	58; 92	58; 90	61; 88	55; 94	55; 94
Percent female [%]	31	61	44	63	51
CFHR5$_{fs}$ variant status					
C/C (non-carrier, n/c)	62	74	73	73	282 (71%)
C/CAA (heterozygote)	19	28	17	21	85 (21%)
CAA/CAA (homozygote)	0	7	11	14	32 (8%)
AMD x *CFHR5*$_{fs}$ variant status					
AMD, *CHFR5*$_{fs}$ n/c	37	49	48	48	182 (46%)
AMD, *CFHR5*$_{fs}$ het	2	11	0	4	17 (4%)
AMD, *CFHR5*$_{fs}$ hom	0	0	0	0	0 (0%)
no AMD, *CHFR5*$_{fs}$ n/c	25	25	25	25	100 (25%)
no AMD, *CFHR5*$_{fs}$ het	17	17	17	17	68 (17%)
no AMD, *CFHR5*$_{fs}$ hom	0	7	11	14	32 (8%)

SD, standard deviation; n/c, non-carrier; het, hterozygote; hom, homozygote

Table 1. Recall study cohort demographics by participating biobank.

We measured the serum levels of 6,627 circulating proteins, including FH and FHR1-5, using the SomaScan platform[27]. One sample failed quality control, leaving a final sample size of 399. Serum levels of FHR-5 were reduced in homozygous and in heterozygous *CFHR5*$_{fs}$ carriers in a dose-dependent manner ($p=4.72x10^{-71}$ after controlling for age at sampling, sex and AMD status), with no residual levels detectable in homozygotes (**Figure 3a**). There was no evidence for batch effects by biobank (**Extended Data Figure 5**). Findings were confirmed with three different FHR-

5 somamers (**Extended Data Figure 6**) and validate that p.Glu163insAA indeed abolishes FHR-5 serum protein levels. Notably, also levels of FHR-2 and FHR-4 were found to be reduced dose-dependently in *CFHR5*$_{fs}$ carriers ($p=9.85x10^{-24}$ and $p=1.23x10^{-7}$, respectively, after controlling for age at sampling, sex and AMD status; **Figure 3a**; **Supplementary Table 10**). Conversely, no differences were found for FH, FHR-1 and FHR-3 ($p>0.05$). FH and FHR1-5 levels were not independently associated with AMD after controlling for *CFHR5*$_{fs}$ carrier status in the recall cohort.

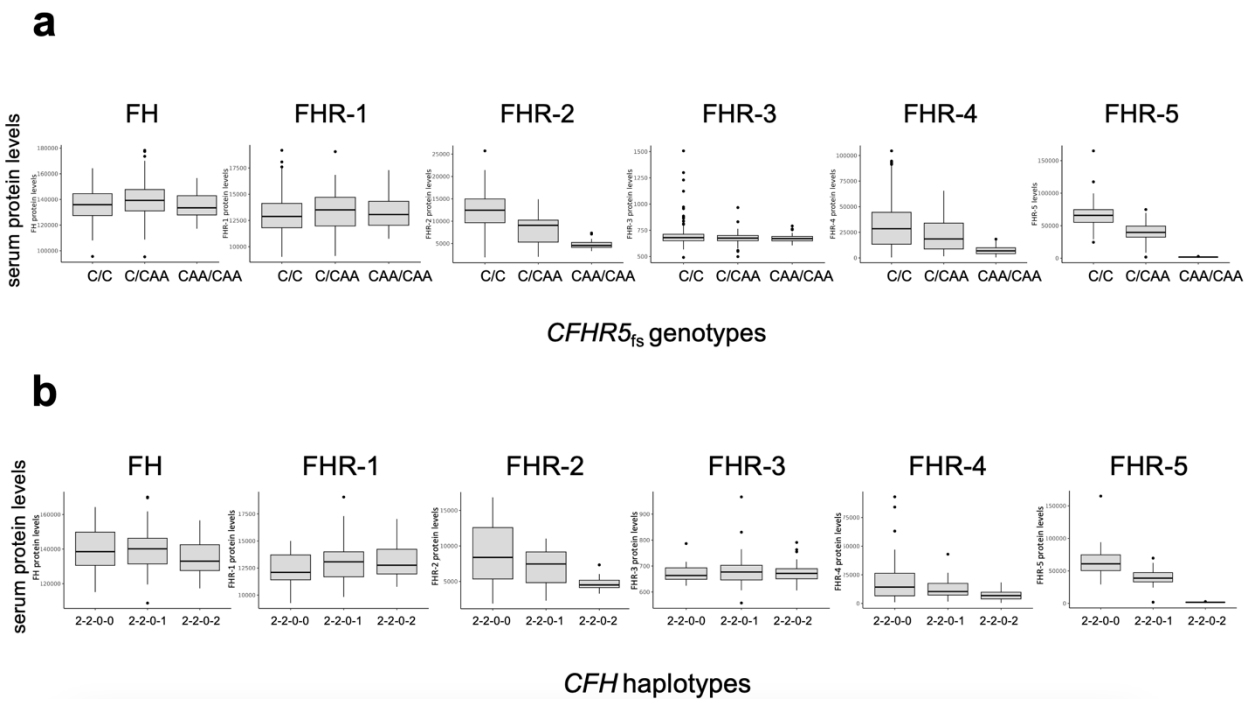


Figure 3. *CFHR5*$_{fs}$ lowers FHR-5, FHR-2 and FHR-4 blood levels in a dose-dependent manner. Relative protein levels of complement factor H (FH) and complement factor H related factors FHR1-5 as measured by representative somamers on the SomaScan platform from serum samples of 399 FinnGen participants recalled from four Finnish biobanks. **(a)** (x-axis) reflects carrier status for *CFHR5* frameshift variant p.Glu163insAA. C/C, non-carriers; C/CAA, heterozygotes; CAA/CAA, homozygotes. **(b)** (x-axis) reflects *CFH* haplotype status. Recall study participants with the highest genetic protection from AMD due to homozygosity for protective alleles at *CFH* rs1410996 and p.Tyr402His (2-2-0-0; n=26) were compared to participants who on top either carry one (2-2-0-1; n=34) or two (2-2-0-2; n=30) copies of *CFHR5*$_{fs}$ alleles. Linear regression analyses were performed. In the box plots, the center line represents the median, the whiskers indicate standard deviations, and the dots indicate highest and lowest values.

We next tested whether the effects of *CFHR5*$_{fs}$ status on serum protein levels synergized with other variation at the *CFH* locus. For this, we grouped the 500,348 genotyped FG participants (DF12) according to haplotypes based on the four regional AMD signals that had remained significant after our stepwise conditional analyses (FG_{AMD}1-4). Five risk groups were defined based on whether individuals carried only risk alleles (0-0-0-0) or were heterozygous (=1) or homozygous (=2) for one or several of the four protective variants in the *CFH* region. As expected, individuals who carried risk alleles for all four signals showed the highest probability to be diagnosed with AMD (OR=1.67), which is reflected by 56.2% of AMD cases in FG falling into this highest risk category (**Figure 4**). Conversely, individuals who carry either *CFHR5* p.Gly278Ser or *CFHR5*$_{fs}$ on top of protective alleles at *CFH* rs1410996 and p.Tyr402His show the highest protection from AMD ($p\_cond=1.5x10^{-18}$, OR=0.59 and $p\_cond=2.7x10^{-9}$, beta=0.52,

respectively). Notably, among the recall study participants, the dose-dependent reduction of FHR-5 ($7.55x10^{-19}$), FHR-2 ($p=1.33x10^{-5}$) and FHR-4 ($2.63x10^{-3}$), but not of FH, FHR-1 and FHR-3 ($p>0.05$) remained significant in individuals who carried one (2-2-0-1; n=34) or two (2-2-0-2; n=30) copies of *CFHR5*$_{fs}$ relative to zero (2-2-0-0; n=26) copies on the most protective haplotype (**Figure 3b, Supplementary Table 11)**. This demonstrates that the protective effects of *CFHR5*$_{fs}$ on serum protein levels are independent of and add to that from other protective variation at the *CFH* locus. A significant reduction in FHR-2, FHR-4 and FHR-5 could also be seen in carriers of *CFHR5* p.Gly278Ser (**Supplementary Table 12**).

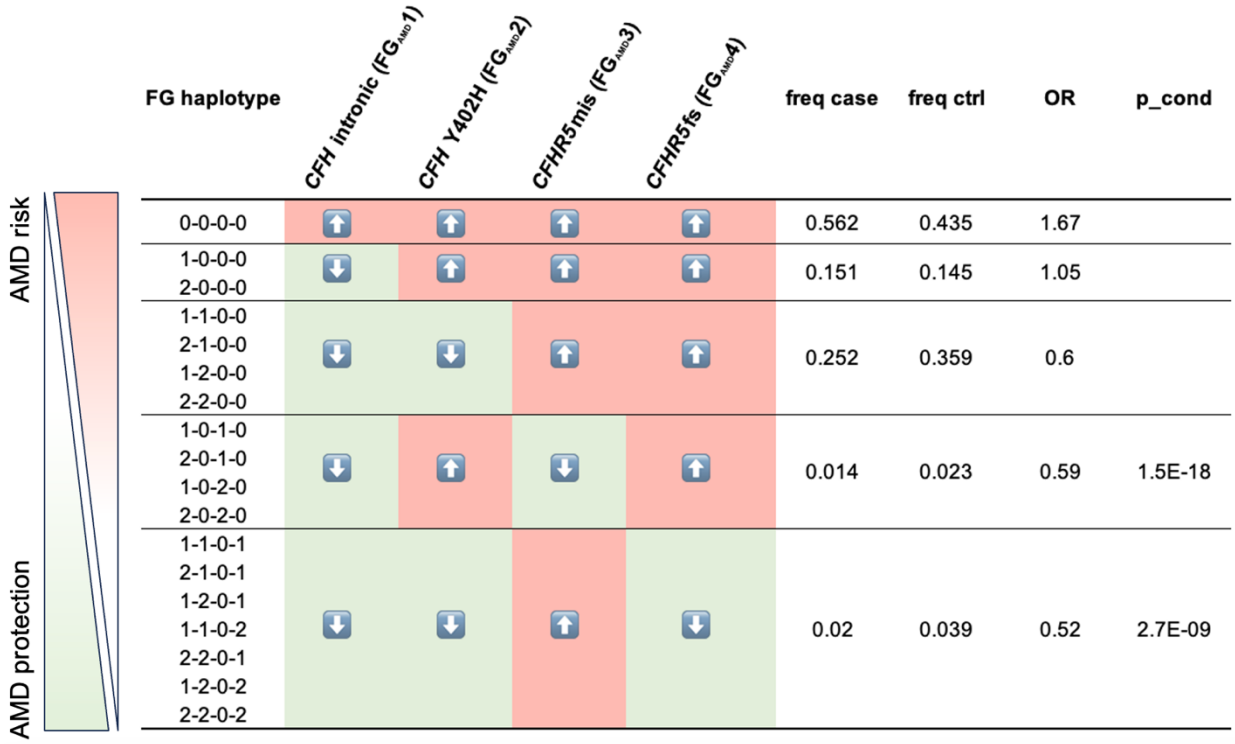


Figure 4. FinnGen *CFH* regional haplotypes and relation to AMD risk in the population. 500,348 FinnGen (FG) participants with available genotypes (DF12) were grouped according to *CFH* regional haplotypes based on the four regional AMD signals that had remained significant after stepwise conditional analyses (FG$_{AMD}$1-4). Five risk groups were defined based on whether individuals carried only risk alleles (0-0-0-0) or were heterozygous (=1) or homozygous (=2) for one or several of the four protective variants in the *CFH* region. OR, odds ratio for association with a diagnostic entry for AMD (H7_AMD) based on FG health records. freq, frequency of diagnostic entry per group. p_cond, p-value after conditioning for FG$_{AMD}$1 (rs1410996) and FG$_{AMD}$2 (p.Tyr402His).

To confirm the robustness and generalizability of the proteomic findings in our recall study, we sought to validate the effect of *CFHR5*$_{fs}$ on reducing FHR-2, FHR-4 and FHR-5 blood levels also in a wider FG cohort not enriched for AMD from which serum samples had been analyzed both, with the Somalogic panel (N=881) as well as an independent proteomics platform (Olink; N=1,732). This yielded highly similar results to our findings in the recall setting (**Supplementary Table 13**). Our results are further consistent with a burden of coding variants in *CFHR5* lowering FHR-5, FHR-2 and FHR-4 plasma levels in UK Biobank (**Supplementary Table 14**)[28], thus replicating that genetic loss-of-function of *CFHR5* reduces FHR-5, FHR-2 and FHR-4 protein levels in blood also in an independent European cohort.

CFHR5 frameshift carriers show increased potential to activate complement pathways.

During normal aging and early stages of AMD, complement activity is critical for the removal of retinal debris. Conversely, if overactive or misdirected, the complement system is a central driver of AMD pathology[6]. FH and FHRs are believed to be critical regulators of alternative pathway (AP) C3 convertase activity which amplifies complement activation via the classical (CP) and MBL-lectin pathways (LP) and converges the complement activation signal on a joint terminal pathway (**Figure 5a**)[29]. To assess whether genetic loss of *CFHR5* function modulates complement activity, we measured functional activities of all three complement pathways in sera of a subset of 40 AMD patients and 44 controls from our FG recall study cohort. Notably, we found that activities of both CP and AP, but not LP were significantly increased in homozygous $CFHR5_{fs}$ carriers relative to non-carriers (p<0.001; **Figure 5b**; **Supplementary Table 15**). Heterozygotes significantly differentiated from non-carriers for both CP and AP (p<0.007) only in AMD cases, but not in unaffected controls and here showed intermediate effects between homozygotes and non-carriers (**Extended Data Figure 7**). As seen with protein levels, CP, and AP activities relative to non-carriers (2-2-0-0; n=8) remained increased in individuals who carried one (2-2-0-1; n=7) or two (2-2-0-2; n=12) copies of $CFHR5_{fs}$ when controlled for the presence of the independently protective *CFH* alleles in the region (p<0.001; **Figure 5c**). In summary, results from these functional analyses demonstrate that genetic loss of *CFHR5* function is independently associated with increased activation capacity of the complement system to confer protection from AMD.

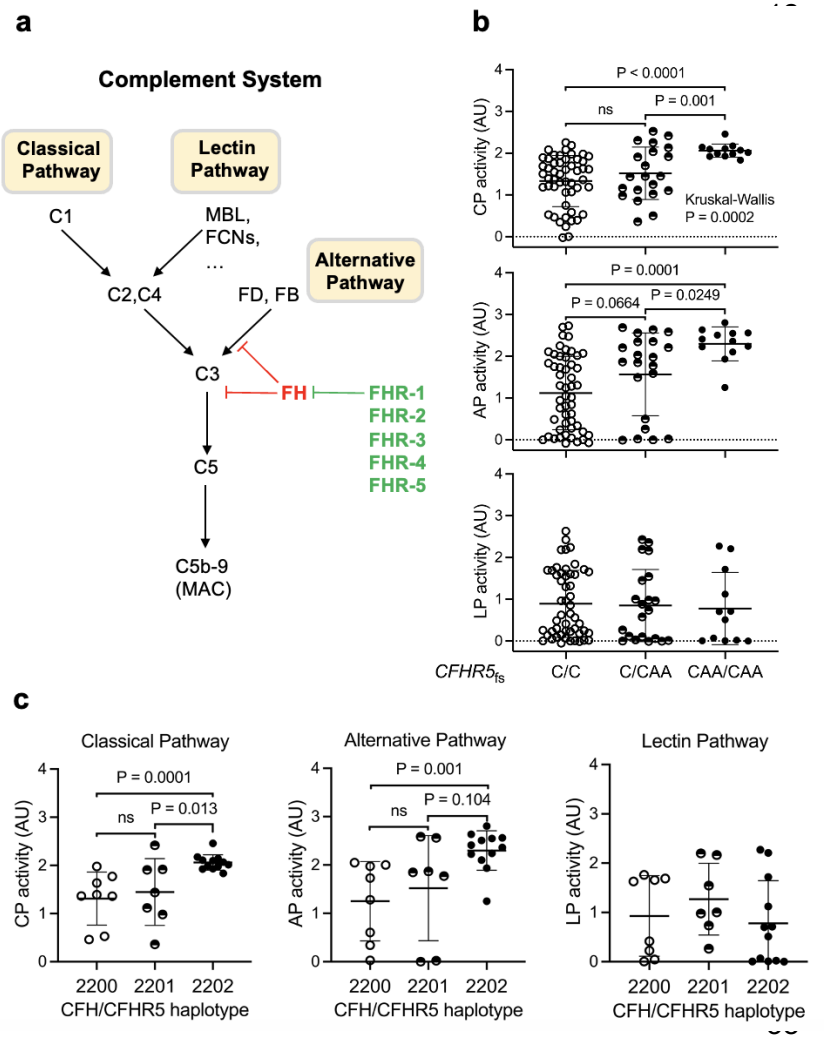


Figure 5. $CFHR5_{fs}$ maintains stronger activity of the classical and alternative complement pathways. (a) High-level sketch of the terminal pathway of the complement system on which the classical pathway (CP), MBL-lectin pathway (LP) and alternative pathway (AP) converge. **(b,c)** Functional analysis of CP, AP, and LP activity in serum samples from 84 recall study participants. (b) (x-axis) reflects carrier status for *CFHR5* frameshift variant p.Glu163insAA. C/C, non-carriers; C/CAA, heterozygotes; CAA/CAA, homozygotes. (c) (x-axis) reflects *CFH* haplotype status for recall study participants. Individuals with an AMD protective haplotype (2-2-0-0; n=8) were compared to individuals who also carried one (2-2-0-1; n=7) or two (2-2-0-2; n=12) copies of $CFHR5_{fs}$ on this haplotype. Statistical analyses were performed with the Student's t-test or the Mann-Whitney U-test depending on whether data were

normally distributed or skewed. The Kruskal-Wallis test was applied for comparisons across all three groups. Each dot represents one individual. For each group the center line represents the median, the whiskers the standard deviation. AU, arbitrary units. For details, see **Supplementary Table 15**.

Discussion

Here, we conducted a GWAS in nearly 475,000 Finns that replicated 17 and identified six novel genomic loci associated with AMD. Using statistical finemapping and conditional analyses, we undertook a fine-grained interrogation of the well-established *CFH* locus. This identified four major *CFH* haplotypes that convey protection from AMD, out of which two are linked to coding variants in the *CFHR5* gene. We demonstrate through a sample recall study that individuals carrying Finn-enriched *CFHR5* loss-of-function variants show a dose-dependent reduction in the *CFHR5* gene product, FHR-5, as well as increased activity of the classical and alternative complement pathways. Our findings establish the relevance of *CFHR5* for AMD risk relative to functionally related genes at the locus and propose therapeutic downregulation of FHR-5 as a promising strategy for the prevention or treatment of AMD.

Our results elucidate the complexity of the first disease locus ever identified through the GWAS approach which at the time required an association study with only 96 AMD patients and 50 controls[9]. Since then, *CFH* has been consistently replicated as an AMD locus, and it has become clear that the *CFH* region is characterized by a substantial heterogeneity across the human population. Notably, it includes several genes with annotated functions in a disease-relevant pathway, large common SVs disrupting some of these genes, and a variable degree of risk between carriers of different regional haplotypes[3,17]. Our study showcases how such complexity can be systematically reduced to pinpoint the most relevant causal contributors to genetic disease risk within a region. For this, we leveraged statistical finemapping in a homogenous population, stepwise conditional analysis of apparently independent highly significant GWAS lead signals, the exclusion of concealed regional variation through sequencing, and customized functional analyses from matched patient and control samples enriched for protective allele carriers in a recall-by-genotype setting. That such a systematic approach to GWAS locus deconvolution is now possible is largely due to the advent of large genetically profiled biobank cohorts such as FG where low-frequency variants enriched in the population enable new discoveries, the complexity of haplotype structure is reduced due to population bottlenecks, and participants are consented *a priori* for dedicated follow-up analyses[23,24].

Consistent with our GWAS findings, multiple lines of evidence have established critical roles for a dysregulated complement system in the etiology of AMD for which an accumulation of lipoproteinaceous drusen in the vicinity of retinal pigment epithelium cells is considered the initial trigger[14]. In 20% of patients this buildup of drusen is followed by a neovascular response, which can be efficiently treated with VEGF inhibitors. In contrast, patients with dry AMD show a slowly progressive loss of retinal pigment epithelium cells through cell death, leading to geographic atrophy and through subsequent deterioration of photoreceptor cells irreversible blindness. While in healthy individuals and early stages of the disease an active complement system is critical for efficient removal of toxic cellular debris, its activation products are also among the most intense mediators of inflammation and can further flare up disease processes. To keep retinal cells healthy, complement activity must therefore be regulated within tight borders. FH enhances the

breakdown of C3 convertase and promotes the cleavage of C3b, thereby finetuning the terminal complement pathway before inflammatory byproducts can be formed. The AMD risk allele p.Tyr402His is believed to reduce interaction between FH and C-reactive protein (CRP) or polyanionic surfaces, leading to a misdirected activation of the terminal complement pathway[13]. Apart from genetics, the central role for FH during AMD pathogenesis has been established through biochemical, cellular, preclinical and clinical studies[30]. FHR proteins, which circulate in blood as dimers and oligomers, have been reported as having antagonistic roles to FH[31]. For instance, complete deletion of *CFHR1* and *CFHR3* through a common SV had initially been thought to be protective of AMD[16]. However, later studies showed that this deletion shares the same haplotype with the *CFH* intronic variant rs1410996 ($FG_{AMD}1$) and its proxies, which are more likely to convey AMD protection[21].

Our new findings establish a critical relevance of FHR5 for complement activity and the risk for AMD. We demonstrate that two Finn-enriched coding variants in *CFHR5*, the frameshift variant p.Glu163insAA ($FG_{AMD}4$) and the missense variant p.Gly278Ser ($FG_{AMD}3$), are associated with a strong, dose-dependent protection from AMD that is unlikely to be explained by other variation in the region. Notably, the protective effect is pronounced even in individuals who already carry a haplotype that conveys a high baseline protection from AMD, which validates *CFHR5* as a causal gene. It further sheds a light on how local additive effects, in this case the co-occurrence of protective alleles in both *CFH* and *CFHR5*, may reduce genetic disease risks even further and augment association signals. Notably, we demonstrate that $CFHR5_{fs}$ carrier status reduces serum levels of the FHR-5 protein, which in homozygous *CFHR5* “knockouts” was undetectable in our assays. The absence of FHR-5 translated into about twofold increased capacity to activate the classical and alternative complement pathways relative to non-carriers. Based on these findings it is tempting to speculate that a higher preserved complement activity may enable carriers of variants lowering FHR-5 levels or function to more efficiently clear cellular debris in their retinas, which consequently would translate into a reduced probability to develop AMD. However, our results also indicate that a mere antagonistic effect might be a too simplistic model to explain the biological interplay between FH and FHR-5. For instance, our recall study from Finnish samples cannot explain whether loss of FHR-5 function would also raise complement activity in a setting where FH is unimpaired, since all carriers of $CFHR5_{fs}$ and *CFHR5* p.Gly278Ser in the FG cohort also carried at least one of the two AMD protective alleles in *CFH*. With the emergence of both sets of variants on the same haplotype, evolution may have found a mechanism to re-adjust the balance between FH and FHR-5 in a most favorable way to prevent disease. Alternatively, while FHR-5 has been found to inhibit binding of FH to C3b at lower concentrations than other FHRs[19], it may not act in isolation, and lower FHR-5 levels could destabilize FHR hetero-oligomers which themselves impact FH and complement activity. In our study, $CFHR5_{fs}$ carriers not only showed reduced FHR-5, but also reduced FHR-2 and FHR-4 levels, with these effects being supported by both FG and UKB, two different proteomics platforms utilized, and for FHR-2 also extended to *CFHR5* p.Gly278Ser carriers. Notably, our fine-grained sequence analysis of the *CFH* region did not provide evidence for yet concealed genetic variation that might explain this observation, so FHR proteins must regulate their mutual abundance on a post-genetic level. Further experiments in future studies are needed to clarify the exact mechanisms how reduction in FHR-5 increases complement activity and lowers AMD risk.

Nevertheless, already now our analyses identify *CFHR5* as an "allelic series" gene[32] that does not have evident associations with on-target safety signals that could be red flags for the development of FHR-5 directed therapies[33]. Treatments against several components of the complement pathway are being explored in clinical trials for their suitability to address especially the dry form of AMD. In 2023, Pegcetacoplan, an intravitreally injected human monoclonal antibody targeting C3[34] was approved for the treatment of geographic atrophy in the US, but not in Europe. A Phase 1 study showed that monthly injections of a recombinant full length human FH protein (GEM103) into the eye are well tolerated[6], with initial results from a Phase 2 study (NCT04643886) showing sustained C3a lowering effects. No therapies have yet targeted *CFHR*1-5 or their products. The option to reduce rather than increase its products to achieve therapeutic benefit makes *CFHR5* a potentially attractive target for neutralizing antibodies or oligonucleotides that, since FHR-5 is synthesized almost exclusively in the liver, could possibly even be explored for systemic administration. That FHR-5 lowering is accompanied by reduction in FHR-2 and FHR-4 indicates that the requirements for a truly isoform-specific modality might not need to be so high. However, like for other complement-targeting therapies, FHR-5 inhibitors might be challenged by a fairly narrow therapeutic range and a need to substantially suppress protein levels. This is supported by our finding that in *CFHR5*$_{fs}$ heterozygotes, preserved complement activation was increased only in cases, but not controls, which suggests that partial reduction in target tissues might not suffice to fully counteract disease processes. Moreover, future studies in longitudinal genetically profiled cohorts will be necessary to better assess whether FHR-5 inhibition might indeed be suited for treatment rather than primarily prevention of AMD.

Our study has several limitations. First, while we powered the recall study sufficiently to demonstrate the effect of *CFHR5*$_{fs}$ on FHR protein levels and complement activation, its size was insufficient for a dissection of how less frequent and even higher resolved *CFH* haplotypes in our cohort impact these processes, which will require larger studies. Also, with its enrichment for AMD cases with available biobank samples our recall study may not be representative for the broader population, although it is reassuring that we could replicate a reduction of FHR proteins in *CFHR5*$_{fs}$ carriers in the subset of the FG population with proteomics data available, as well as in UK Biobank. Notably, the distribution of *CFH* haplotypes in the recall study cohort had a more pronounced effect than AMD case-control status, which may be an important learning for the design of future such studies. Future studies in non-European cohorts with alternative haplotype compositions may also help to further disambiguate protective effects from *CFH* versus *CFHR5*. Our discovery of a single 1KG participant of African ancestry who carries *CFHR5*$_{fs}$ on a different haplotype, as well as of the non-Finnish *CFHR5* frameshift variant p.Glu163insA, indicate that such informative individuals exist. A further limitation is our cross-sectional study design. More in depth analyses in individual-level Finnish health records over an extended period of time, as well as obtaining additional samples and clinical data, such as retinal scans, from informative *CFHR5*$_{fs}$ carriers could help to prioritize the optimal target population for testing *CFHR5*-directed therapies.

In summary, our study exemplifies how existing and newly acquired data from a broadly consented human population cohort like FG can be leveraged and combined to inform genetic and post-genetic research well beyond the discovery of novel genetic leads. With the wealth of

information captured, future research should enable a more granular understanding on how links between distinct haplotypes at GWAS loci are influenced by concomitant genetic variation elsewhere in the genome (such as at the similarly prominent *ARMS2/HTRA1* locus), environmental contributors (such a smoking, which further increases a genetically elevated risk for AMD in *CFH* p.Tyr402His carriers[35]), or aging. It should further facilitate a more customized design of research cohorts for clinical trials, for instance to enrich a trial population for individuals with the most informative genetic profiles to demonstrate efficacy. With permissive legislation for secondary use of research data, we predict that resources where genetic and deep longitudinal health information are linked at scale will become invaluable tools not only to better understand biology, but also accelerate the path towards efficient novel medicines.

Methods

FinnGen Samples and Participants

FinnGen (FG) (https://www.finngen.fi/en) is a public-private partnership project that aims to generate medically and therapeutically relevant insights into human disease by combining genome and digital health data from over 500,000 Finns. FG participants provide informed consent for biobank research (see below), which includes secondary use of research samples stored at nine participating Finnish biobanks. Individual-level genotypes, register data and an overview of available samples from FG participants can be accessed by approved researchers via the Fingenious portal (https://site.fingenious.fi/en/) hosted by the Finnish Biobank Cooperative FinBB (https://finbb.fi/en/). All analyses in this manuscript rely on FG data freeze (DF) 12, which is anticipated to become public in late 2024.

Genotype Data Quality Control, Association Testing and Conditional Analyses

Array-based genotype data in FG were called and subjected to variant and sample level quality control (QC) followed by phasing and imputation as described in Kurki et al 2023[24]. In FG DF12, this project-wide process resulted in a total of 500,348 individuals after removal of related individuals and non-Finnish ancestry participants. This dataset was used in all FG-wide analyses in this study. GWAS analysis was conducted for phenotype H7_AMD (https://risteys.finregistry.fi/endpoints/H7_AMD), which includes individuals with at least one registry entry of either wet or dry AMD or both, comprising 12,495 cases and 461,686 controls that all passed genotype QC and inclusion criteria. GWAS analysis was performed with REGENIE 2.2.4[36] using a logistic mixed model adjusted for age, sex, genotyping batch and the first ten principal components of ancestry with an approximate Firth test for robust effect size estimation. Plink and vcf files for imputed genotypes, as well as ocular variables from the biobanks were securely transferred and stored on the DNAnexus platform.

Statistical finemapping was conducted using SuSIE[37]. Stepwise conditional analysis was performed identically with the serial addition of established significant variants at the *CFH* locus by adding these variants as covariates. Since associations at the *CFH* locus are highly significant for both wet and dry AMD, we utilized the combined group with any AMD diagnosis (H7_AMD) as the phenotype also for conditional analyses. Phenome-wide association analyses for $CFHR5_{fs}$ (p.Glu163insAA; chr1:196994128:C:CAA, rs565457964) variant carriers were conducted against a total 2,405 FG DF12 phenotypes accessible via the PheWeb browser (https://r12.finngen.fi) and

displayed in a LAVAA plot (https://geneviz.aalto.fi/LAVAA/) (**Extended Data Figure 2**) as described in Fauman et al 2023[38].

Two previously reported structural variants (SVs) within the *CFH* region were called from Finnish participants in gnomAD version 4 (https://gnomad.broadinstitute.org) from an SV map compiled from whole genome sequencing (WGS) data. Proxy SNPs were identified to impute these SVs into the broader FG population and association testing for H7_AMD was conducted as described above.

CFH Region Genetic Architecture in 1000 Genomes cohort

We used phased short-read and long-read sequencing data from the expanded 1000 Genomes (1KG) project[25,26] as well as assembled haplotypes from the Human Pangenome Reference Consortium (HPRC) (https://humanpangenome.org)[39] to catalog haplotype-resolved genetic variation across the *CFH* region (**Supplementary Table 6**). For the HPRC haplotypes, we used AGC[39] to extract the haplotype sequences from the AGC archive of HPRC year1 assemblies, which contains fully phased diploid assemblies for 47 samples. We then mapped the GRCh38 *CFH* region sequence to each assembly using minimap2[40] with the option -x asm10, extracted the mapped subsequence using samtools[41] and then aligned all HPRC *CFH* regional haplotypes to GRCh38. We visualized all alignments using dot plots[42] and IGV[43] to reveal divergent genetic architectures present in the haplotypes (**Extended Data Figure 3**).

To call *CFH* haplotypes, we extracted eight GWAS lead SNPs from the 1KG long-read data that had been utilized previously to construct haplotypes conveying different degrees of AMD risk[17]. We further refined these haplotypes by also considering 16 additional variants that now included regional SVs with allele frequency >1% across the 1KG cohort and the GWAS lead signals from our FG DF12 AMD GWAS (**Supplementary Table 5**). This identified 254 different haplotypes in the *CFH* region across the 6,406 chromosomes of all subpopulations. Out of these, only three showed a frequency of over 15% across all populations, 12 of >1% and 42 of >0.1%. 214 of the 254 haplotypes were present in <10 copies, with nearly 50% being singletons (**Supplementary Table 6**).

To estimate *CFH* haplotype frequencies across ancestries, we leveraged phased short-read data from the 1KG cohort[25] available at the International Genome Sample Resource (IGSR). We first subset the phased variants to the *CFH* region and the 2,504 unrelated samples (5,008 haplotypes) and then aggregated shared haplotypes by self-identified ancestries from 26 geographic locations in five continental regions (labeled as populations and super-populations for consistency with prior 1KG literature). The 198 chromosomes from unrelated Finns in the 1KG cohort distributed across 34 haplotypes, out of which 93% were explained by the three most common haplotypes overall and the remainder being derivatives of these. The allele count of each *CFH* regional haplotype per ancestry is shown in **Supplementary Table 7**.

We also queried new, long-read 1KG data available for a subset of 1,019 samples[26] to identify potentially missed SVs associated with the *CFHR5* frameshift variants p.Glu163insAA and p.Glu163insA, since short-reads have limitations in identifying SVs in repeat-rich and complex regions of the genome. Out of the 1,019 samples, one sample was a heterozygous carrier for p.Glu163insA (NA20127) and four samples were heterozygous carriers for p.Glu163insAA (HG00268, NA20342, NA20531, NA20813). For all 5 samples, we did not identify additional structural variants with the long-read data within the *CFH* region, except for an intronic 86bp

deletion in four homozygous and one heterozygous carrier that resides within a tandem repeat region and shows an allele frequency of over 80% across the cohort. In summary, based on long-read data for 1,019 samples, we found no evidence for cryptic regional variation that could explain the *CFHR5*$_{fs}$ effects.

Recall Study Cohort Selection

For most of the 500,348 genotyped FG participants, samples are stored at and can be retrieved from nine regional biobanks across Finland. We identified the FG subset with serum samples available in four biobanks, Auria, Borealis, Eastern Finland, and Tampere Biobank. Of these, we excluded samples with low quality genotypes at key sites (imputation info score <0.95) and individuals with highly unlikely *CFH* haplotypes suggestive of potential genotyping errors. Of the remainder, 200 AMD cases and 200 controls were selected for a sample recall study based on individual-level phenotype data deposited at the respective biobanks. FG participants were defined as cases if they had at least one registry entry with ICD codes for dry AMD (H35.30, 2625A) at the time of the study, as well as none of the following: ICD codes for wet AMD, hereditary conditions affecting the eye (e.g., Stargardt disease, hereditary retinal dystrophy), macular hole, macular pucker, diabetic retinopathy, glaucoma, vitreous hemorrhage, separation of retinal layers, onset of blindness at <30 years of age, or injections into the eye). Controls were required to have no ICD entries for AMD, be 65 years of age or older at time of sampling and were matched to cases based on age and sex. To maximize our ability to examine the effect of the *CFHR5*$_{fs}$ variant on serum protein levels and complement activity, we originally aimed at an approximately equal distribution of *CFHR5*$_{fs}$ carriers and non-carriers between both groups. However, given an allele frequency of this frameshift variant of ~4% in the overall Finnish population together with its strong AMD protective effect, we found the allele substantially underrepresented in AMD cases. Thus, the final recall study cohort (n=400) included only 17 heterozygous cases (as opposed to 68 heterozygous controls) while no samples from a very small number of homozygous FG participants with diagnostic codes for dry AMD could be retrieved (as opposed to 32 homozygous controls).

Proteomics Data Acquisition

200µl of serum from each recall study participant were thawed, aliquoted into tubes of 130µl for proteomic profiling and shipped on dry ice from Finland to the Somalogic lab in Boulder, CO, USA. All samples underwent the Somalogic SomaScan assay to measure expression of 7,000 proteins including FH and FHR1-5. Somalogic generated a sample quality report which flagged three of the 400 samples. Of these, only one sample was far outside of the normal QC range and hence excluded from analysis, leaving a final sample size of 399. The majority of values for FH and FHR1-5 were within the 95% normal serum RFU range provided by Somalogic, with the exception of an extreme outlier for FHR-3, which we excluded. FHR-1 and FHR-5 both were targeted by multiple somamers. One of the FHR-1 somamers (15468-14) showed evidence of binding to FH with similar affinity, thus we used the other FHR-1 somamer (5982-50) in our analyses. All three FHR-5 somamers showed similar QC metrics and results, so for simplicity we focused on a the somamer with the lowest serum %CV (3666-17) in our main analyses, with all three FHR-5 somamers showing similar effects on FHR-5, FHR-4 and FHR-2 levels (**Extended Data Figure 6**). Statistical analyses from proteomics data were performed on the DNAnexus platform using R

as described previously[26,44]. Linear regression analyses were used to determine association between genotypes and protein levels, and between AMD and protein levels.
Analyses on how carrier status for AMD-protective alleles in *CFH* and *CHFR5* impacted levels of FH and FHR1-5 in plasma of 881 FG participants profiled with the Somalogic panel, and in 1,732 FG participants profiled with the Olink panel were conducted in the FinnGen Sandbox. Results are provided as **Supplementary Table 13**. To identify plasma proteins associated with the burden of protein-coding variants in *CFH* and *CFHR*1-5 genes we relied on results from the UK Biobank Pharma Proteomics Project in which 49,736 participants of European ethnicity had been profiled with the Olink panel as reported in Dhindsa et al 2023[27].

Complement Activity Measurements:

70µl of serum from each recall study participant were shipped on dry ice for complement activity profiling to the Finnish Institute for Molecular Medicine (FIMM) in Helsinki, Finland. Functional activities of the classical, alternative and MBL-lectin complement pathways were measured using a commercial enzyme immunoassay (WIESLAB® Complement System Screen, Cat# COMPL300). This assay allows to determine activities of all three pathways by measuring the amount of the terminal complement complex component C5b-9, which is produced as a result of complement activation. The amount of C5b-9 generated is proportional to the functional activity of complement pathways. Statistical analyses of differences between *CHFR5*$_{fs}$ non-carriers, heterozygotes, and homozygotes across the entire subcohort profiled for complement activity (n=84) or adjusted for AMD status, as well as recall study participants with different degrees of protection from Finnish AMD protective haplotypes (n=27) were performed in Graphpad Prism according to the distribution of data. Student's t-test was used for normally distributed data and Mann-Whitney U-test for skewed distributed data.

FinnGen (DF12) Ethics Statement

FG study participants provided informed consent for biobank research based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected before the Finnish Biobank Act came into effect (in September 2013) and the start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.
The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020, THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021,

THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/2021, Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20) TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4th July 2019.

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 12 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, BB2021_65, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, HUS/430/2021 §28, §29, HUS/150/2022 §12, §13, §14, §15, §16, §17, §18, §23, §58, §59, HUS/128/2023 §18, Auria Biobank AB17-5154 and amendment #1 (August 17 2020) and amendments BB_2021-0140, BB_2021-0156 (August 26 2021, Feb 2 2022), BB_2021-0169, BB_2021-0179, BB_2021-0161, AB20-5926 and amendment #1 (April 23 2020) and it´s modifications (Sep 22 2021), BB_2022-0262, BB_2022-0256, Biobank Borealis of Northern Finland_2017_1013, 2021_5010, 2021_5010 Amendment, 2021_5018, 2021_5018 Amendment, 2021_5015, 2021_5015 Amendment, 2021_5015 Amendment_2, 2021_5023, 2021_5023 Amendment, 2021_5023 Amendment_2, 2021_5017, 2021_5017 Amendment, 2022_6001, 2022_6001 Amendment, 2022_6006 Amendment, 2022_6006 Amendment, 2022_6006 Amendment_2, BB22-0067, 2022_0262, 2022_0262 Amendment, Biobank of Eastern Finland 1186/2018 and amendment 22§/2020, 53§/2021, 13§/2022, 14§/2022, 15§/2022, 27§/2022, 28§/2022, 29§/2022, 33§/2022, 35§/2022, 36§/2022, 37§/2022, 39§/2022, 7§/2023, 32§/2023, 33§/2023, 34§/2023, 35§/2023, 36§/2023, 37§/2023, 38§/2023, 39§/2023, 40§/2023, 41§/2023, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), BB2021-0140 8§/2021, 9§/2021, §9/2022, §10/2022, §12/2022, 13§/2022, §20/2022, §21/2022, §22/2022, §23/2022, 28§/2022, 29§/2022, 30§/2022, 31§/2022, 32§/2022, 38§/2022, 40§/2022, 42§/2022, 1§/2023, Central Finland Biobank 1-2017, BB_2021-0161, BB_2021-0169, BB_2021-0179, BB_2021-0170, BB_2022-0256, BB_2022-0262, BB22-0067, Decision allowing to continue data processing until 31st Aug 2024 for projects: BB_2021-0179, BB22-0067,BB_2022-0262, BB_2021-0170, BB_2021-0164, BB_2021-0161, and BB_2021-0169, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020, Finnish Hematological Registry and Clinical Biobank decision 18th June 2021, Arctic biobank P0844: ARC_2021_1001.

Data availability

FG summary association results are being released bi-annually via https://www.finngen.fi/en/access_results and can be explored in a public results browser (at time of submission: https://r11.finngen.fi). All analyses in this manuscript which rely on variants that were directly interrogated through chip-based genotyping with the FG array or imputed rely on FG data freeze 12, which is anticipated to become public in late 2024. Individual-level genotypes and register data from FG participants can be accessed by approved researchers via the Fingenious portal (https://site.fingenious.fi/en/) hosted by the Finnish Biobank Cooperative FinBB (https://finbb.fi/en/). Data release to FinBB is timed to the bi-annual public release of FG summary results which occurs twelve months after FG consortium members can start working with the data.

Code availability

FG data analysis pipelines are freely available from https://github.com/FINNGEN/. The FinnGen Handbook, https://finngen.gitbook.io/documentation/, contains a detailed description of data production and analysis, including code used to run the analyses described in this manuscript.

References

1. Rudnicka, A. R. et al. Age and gender variations in age-related macular degeneration prevalence in populations of European ancestry: a meta-analysis. Ophthalmology 119, 571–580 (2012).

2. Seddon, J. M., George, S., Rosner, B. & Klein, M. L. CFH gene variant, Y402H, and smoking, body mass index, environmental associations with advanced age-related macular degeneration. Hum. Hered. 61, 157–165 (2006).

3. Fritsche, L. G. et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat. Genet. 48, 134–143 (2016).

4. He, W. et al. Genome-Wide Meta-analysis Identifies Risk Loci and Improves Disease Prediction of Age-Related Macular Degeneration. Ophthalmology 131, 16–29 (2024).

5. Gormann, B. R. et al. Distinctive cross-ancestry genetic architecture for age-related macular degeneration. medRxiv (2022) doi.org/10.1101/2022.08.16.22278855.

6. Khanani, A. M. et al. A Phase I, Single Ascending Dose Study of GEM103 (Recombinant Human Complement Factor H) in Patients with Geographic Atrophy. Ophthalmol Sci 2, 100154 (2022).

7. Hageman, G. S. et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. Proc. Natl. Acad. Sci. U. S. A. 102, 7227–7232 (2005).

8. Haines, J. L. et al. Complement factor H variant increases the risk of age-related macular degeneration. Science 308, 419–421 (2005).

9. Klein, R. J. et al. Complement factor H polymorphism in age-related macular degeneration. Science 308, 385–389 (2005).

10. Edwards, A. O. et al. Complement factor H polymorphism and age-related macular degeneration. Science 308, 421–424 (2005).

11. Maller, J. et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nat. Genet. 38, 1055–1059 (2006).

12. Li, M. et al. CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. Nat. Genet. 38, 1049–1054 (2006).

13. Laine, M. et al. Y402H polymorphism of complement factor H affects binding affinity to C-reactive protein. J. Immunol. 178, 3831-3836 (2007).

14. Kijlstra, A. & Berendschot, T. T. J. M. Age-related macular degeneration: a complementopathy? Ophthalmic Res. 54, 64–73 (2015).

15. Raychaudhuri, S. et al. A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. Nat. Genet. 43, 1232–1236 (2011).

16. Hughes, A. E. et al. A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. Nat. Genet. 38, 1173–1177 (2006).

17. Lorés-Motta, L. et al. Common haplotypes at the CFH locus and low-frequency variants in CFHR2 and CFHR5 associate with systemic FHR concentrations and age-related macular degeneration. Am. J. Hum. Genet. 108, 1367–1384 (2021).

18. Närkiö-Mäkelä, M., Hellwage, J., Tahokallio, O. & Meri, S. Complement-regulator factor H and related proteins in otitis media with effusion. Clin. Immunol. 100, 118-126 (2001).

19. Goicoechea de Jorge, E. et al. Dimerization of complement factor H-related proteins modulates complement activation in vivo. Proc. Natl. Acad. Sci. U. S. A. 110, 4685–4690 (2013).

20. Cantsilieris, S. et al. Recurrent structural variation, clustered sites of selection, and disease risk for the complement factor H () gene family. Proc. Natl. Acad. Sci. U. S. A. 115, E4433–E4442 (2018).

21. Raychaudhuri, S. et al. Associations of CFHR1-CFHR3 deletion and a CFH SNP to age-related macular degeneration are not independent. Nat. Genet. 42, 553–5; author reply 555–6 (2010).

22. Zouache, M. A. et al. Levels of complement factor H-related 4 protein do not influence susceptibility to age-related macular degeneration or its course of progression. Nat. Commun. 15, 443 (2024).

23. Sun, B. B. et al. Genetic associations of protein-coding variants in human disease. Nature 603, 95–102 (2022).

24. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 613, 508–518 (2023).

25. Byrska-Bishop, M. et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell 185, 3426–3440.e19 (2022).

26. Schloissnig, S. et al. Long-read sequencing and structural variant characterization in 1,019 samples from the 1000 Genomes Project. bioRxiv (2024) doi:10.1101/2024.04.18.590093.

27. Sun, B. B. et al. Genomic atlas of the human plasma proteome. Nature 558, 73–79 (2018).

28. Dhindsa, R. S. et al. Rare variant associations with plasma protein levels in the UK Biobank. Nature 622, 339–347 (2023).

29. Meri, S. Self-nonself discrimination by the complement system. FEBS Lett. 590, 2418–2434 (2016).

30. den Hollander, A. I. et al. Systems genomics in age-related macular degeneration. Exp. Eye Res. 225, 109248 (2022).

31. Hebecker, M. & Józsi, M. Factor H-related protein 4 activates complement by serving as a platform for the assembly of alternative pathway C3 convertase via its interaction with C3b protein. J. Biol. Chem. 287, 19528–19536 (2012).

32. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. Nat. Rev. Drug Discov. 12, 581–594 (2013).

33. Diogo, D. et al. Phenome-wide association studies across large population cohorts support drug target validation. Nat. Commun. 9, 4285 (2018).

34. Heier, J. S. et al. Pegcetacoplan for the treatment of geographic atrophy secondary to age-related macular degeneration (OAKS and DERBY): two multicentre, randomised, double-masked, sham-controlled, phase 3 trials. Lancet 402, 1434–1448 (2023).

35. Seddon, J. M., Cote, J., Page, W. F., Aggen, S. H. & Neale, M. C. The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences. Arch. Ophthalmol. 123, 321–327 (2005).

36. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat. Genet. 53, 1097–1103 (2021).

37. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the 'Sum of Single Effects' model. PLoS Genet. 18, e1010299 (2022).

38. Fauman, E. B. et al. LAVAA: a lightweight association viewer across ailments. Bioinform Adv 3, vbad018 (2023).

39. Deorowicz, S., Danek, A. & Li, H. AGC: compact representation of assembled genomes with fast queries and updates. Bioinformatics 39, (2023).

40. Li, H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100 (2018).

41. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

42. Rausch, T. et al. Long-read sequencing of diagnosis and post-therapy medulloblastoma reveals complex rearrangement patterns and epigenetic signatures. Cell Genom 3, 100281 (2023).

43. Robinson, J. T. et al. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26 (2011).

44. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature 622, 329–338 (2023).

Acknowledgements

We thank all the participants, contributors, and researchers of FinnGen and its participating biobanks for making data or samples available for this study. We are grateful to Marcel Messing for help with conducting complement analyses. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta), Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/) and Arctic Biobank (https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank). All Finnish Biobanks are members of BBMRI.fi infrastructure (https://www.bbmri-eric.eu/national-nodes/finland/). Finnish Biobank Cooperative -FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (https://site.fingenious.fi/en/) managed

by FINBB. Further support for the study came from Biogen Inc, the Finnish Institute of Molecular Medicine (FIMM), the Academy of Finland (Project #336411), the Sigrid Jusélius Foundation (#4708373), Special State Subsidy for Health Research at Helsinki University Hospital (VTR-funding, TYH2023322), the National Institutes of Health grants MH115957 and HG011450, and the Health + Life Science Alliance Heidelberg Mannheim.

Author contributions

Conceptualization and design: MPR, SL, MD, HR. Methodology: MPR, SL, EN, TR, SM, MD, HR. Analysis: MPR, SL, EN, TR, ZZ, PBP, SDI, EA, EC, YO, MK, MD, HR. Experimental work: EN, EA, EC, SM. FinnGen/biobank protocols and analysis: MPR, HML, JM, MD, HR, FinnGen. Supervision and/or Funding: MK, MT, JK, KC, SM, MD, HR, FinnGen. Writing: MPR, SL, MD, HR. All authors critically reviewed the manuscript.

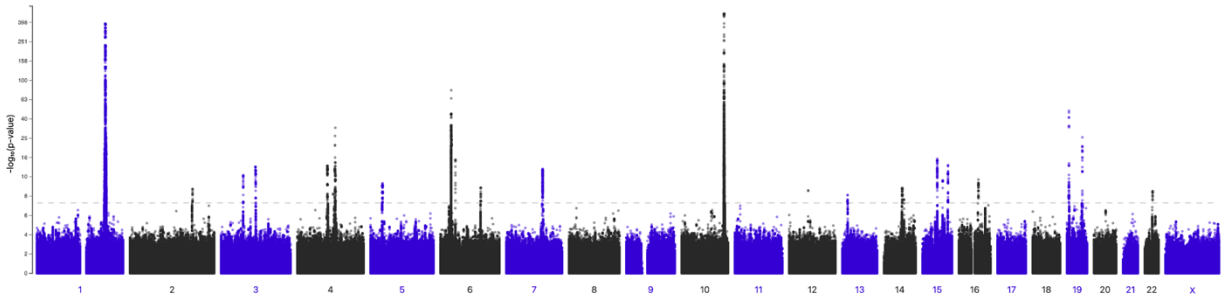
Competing interests

SL, YO, HML, KC and HR were employees at Biogen during data generation for this study. SL is an employee of Bristol-Myers Squibb. YO and KC are employees of Johnson & Johnson. HML is an employee of Moderna. JM is an employee of FinBB. MD is a co-founder of Maze Therapeutics. HR is an employee at insitro Inc. The other authors declare no competing interests.

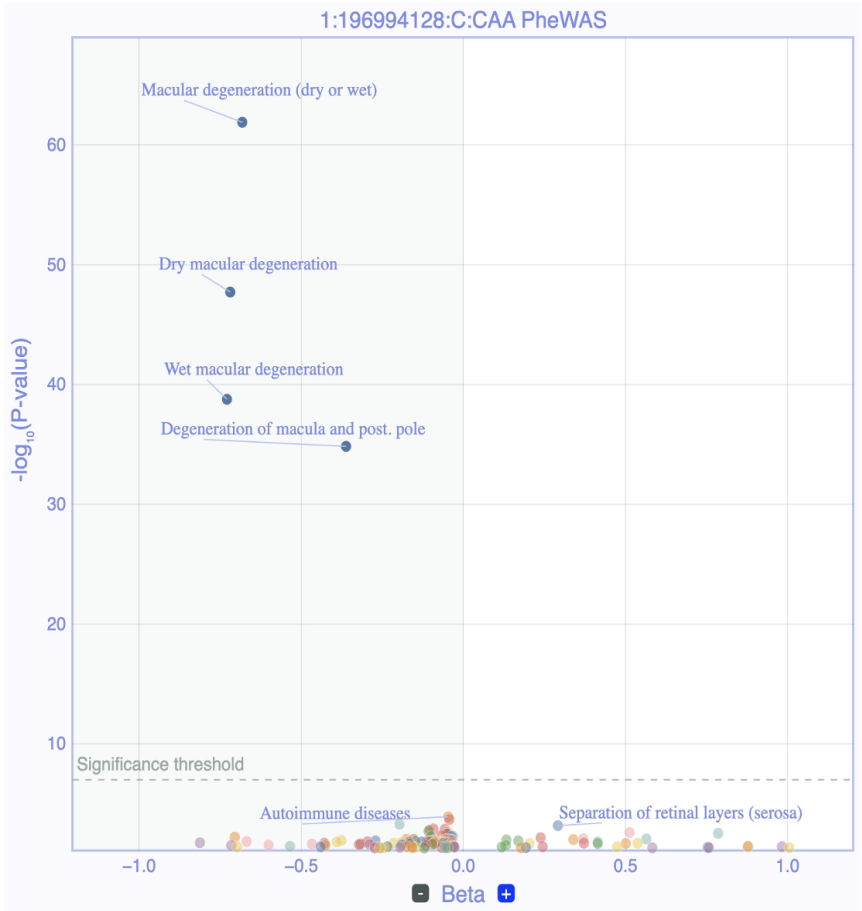
Additional information

This manuscript contains a Supplementary Information file with **Extended Data Figures 1-7** and **Supplementary Tables 1-16**. For a full list of FinnGen authors and their affiliations, see **Supplementary Table 16**.

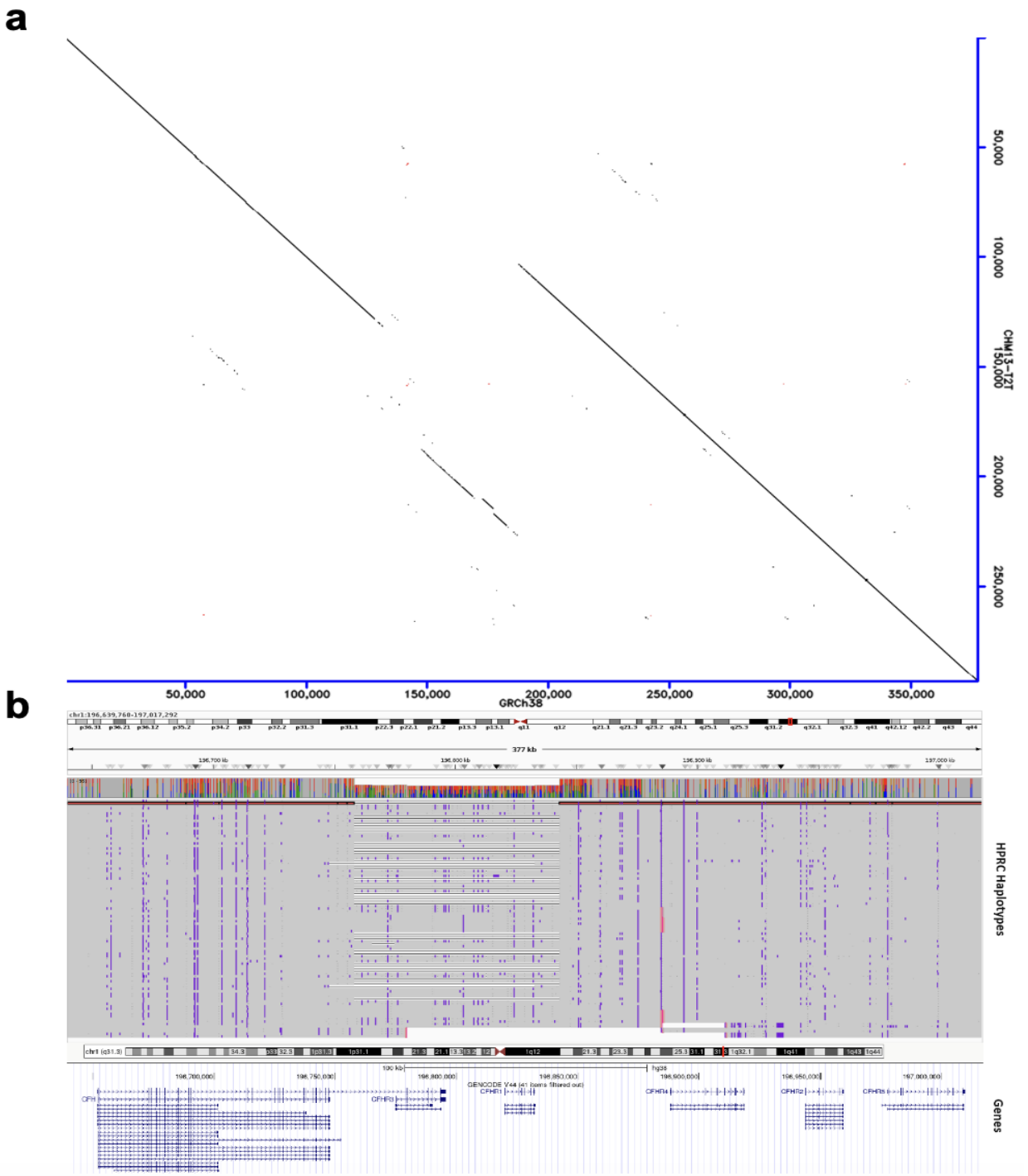
Extended Data Figures



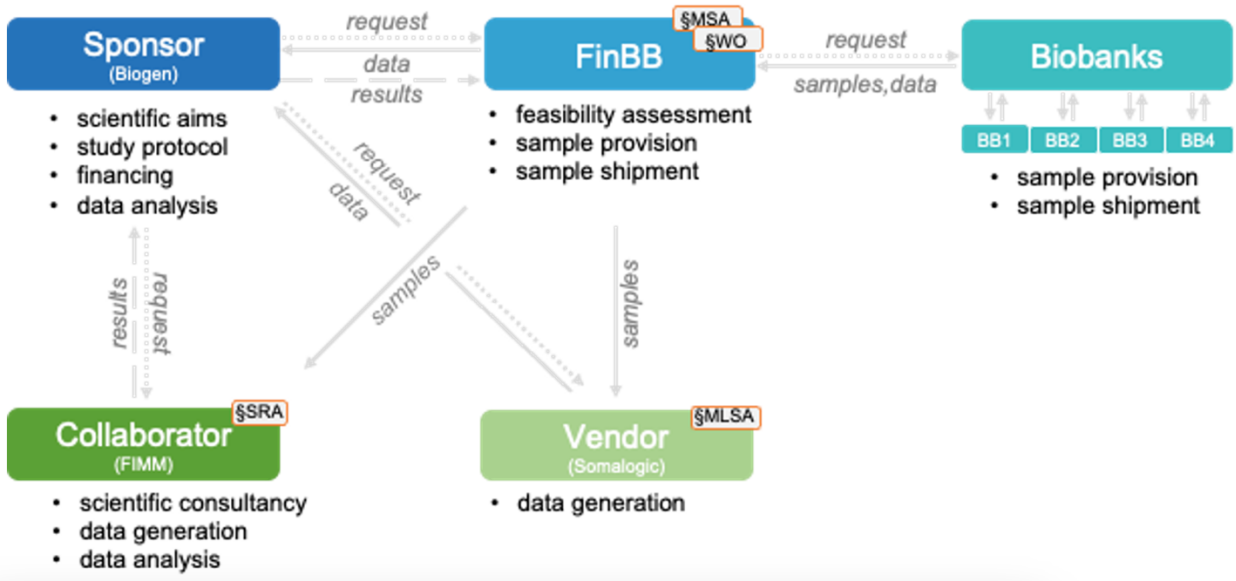
Extended Data Figure 1. Manhattan plot of genome-wide association analysis for age-related macular degeneration. GWS summary statistics are derived from 12,495 cases and 461,686 controls in FinnGen (DF12). AMD cases are defined has having at least one entry with the diagnostic code H7_AMD, reflecting both wet and dry AMD. (x-axis), genomic position of genetic variant; (y-axis), -log10(P-value) of tested genetic variant. Dashed horizontal line indicates genome-wide significance threshold ($P<5x10^{-8}$).



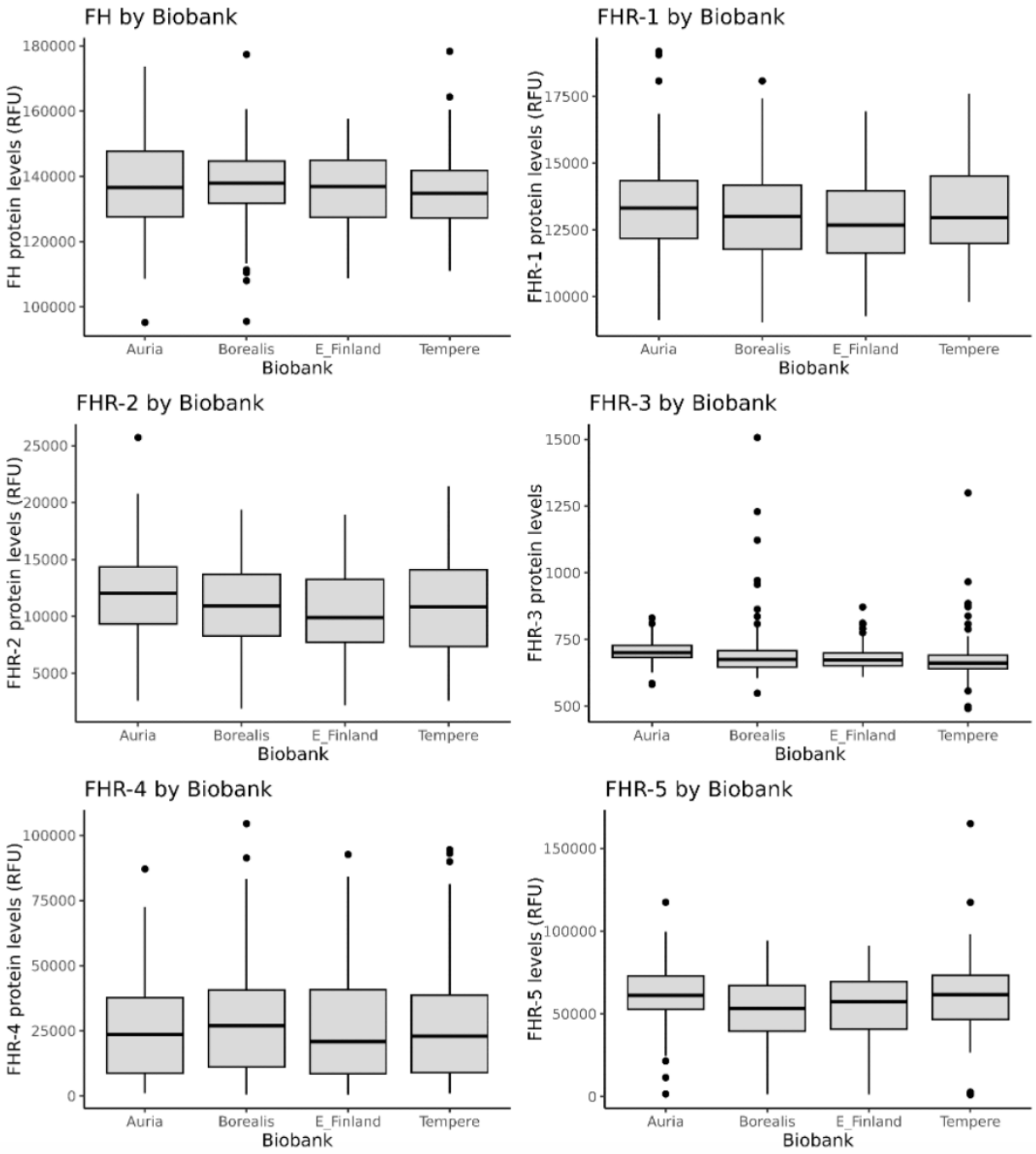
Extended Data Figure 2. Phenome-scan of *CFHR5*$_{fs}$ frameshift variant. LAVAA plot for the Finnish-enriched *CFHR5* frameshift variant p.Glu163insAA (chr1:196994128:C:CAA, rs565457964) in FinnGen (DF12) against 2,408 FinnGen phenotypes. Each dot represents association signals for one respective phenotype. Phenotypes with associations meeting genome-wide significance and of putative relevance for drug repositioning and safety are named. (x-axis), beta of association signal; (y-axis), -log10(P-value) of tested genetic variant. Dashed horizontal line reflects genome-wide significance threshold ($P<5x10^{-8}$).



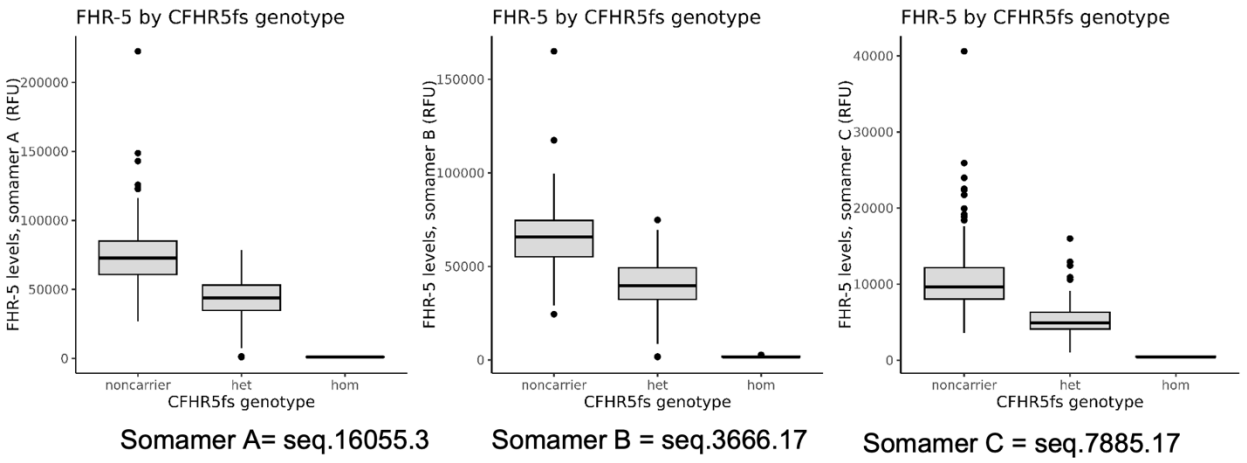
Extended Data Figure 3. Haplotype structure of the *CFH* region as mapped through phased long-read assemblies. (a) Dot plot of the chromosomal region chr1:196,639,760-197,017,292 in GRCh38 assembly (x-axis) compared to the corresponding region chr1:195,986,103-196,279,118 in CHM13-T2T assembly (y-axis). The center of the region reveals a previously described prevalent deletion of ~91kb (HGSV_15069) present in 27.2% of sequenced individuals. **(b)** IGV view of the same region with all Human Pangenome Reference Consortium (HPRC) haplotypes (48 samples) aligned to GRCh38. The CHM13-T2T haplotype is included in this view and highlighted in red. Annotation of protein-coding gene regions reveals that the deletion encompasses *CFHR3* and *CFHR1* genes.



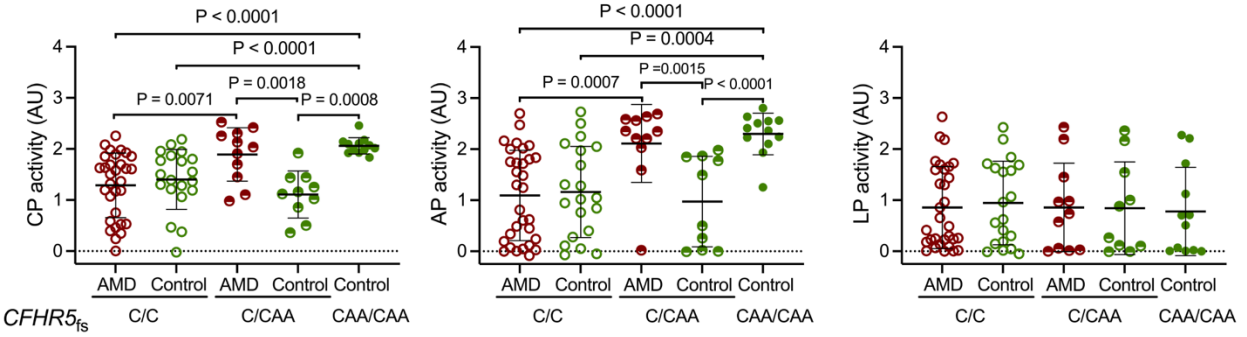
Extended Data Figure 4. Workflow for FinnGen sample recall study from Finnish biobanks. Scientific project definition and drafting of a concrete study protocol were facilitated by the Finnish Biobank Cooperative (FinBB) who conducted feasibility assessments in four selected Finnish biobanks. Available samples from cases and controls were matched based on electronic health data according to the sponsor's in- and exclusion criteria and shipped to a vendor (Somalogic) and academic collaborators at the Finnish Institute for Molecular Medicine (FIMM) for data generation. Data was jointly analyzed by the study's sponsor and academic collaborators. Study realization required independent contracting with FinBB (MSA, master services agreement; WO, work order), the vendor (MLSA, master license and services agreement) and the academic partner (SRA, sponsored research agreement).



Extended Data Figure 5. Serum levels of FH and FHR1-5 proteins by biobank. Relative protein levels of Complement factor H (FH) and Complement factor H related factors FHR1-5 as measured by representative somamers on the SomaScan platform from serum samples of 399 FinnGen participants recalled from four Finnish biobanks. (x-axis) reflects participating biobank. Univariate linear regression analyses were performed using R. In the box plots, the center line represents the median, the box limits represent the interquartile range (IQR) and the whiskers indicate the minimum and maximum values. E_Finland, Biobank of Eastern Finland. RFU, relative fluorescence units.



Extended Data Figure 6. FHR-5 levels by *CFHR5*$_{fs}$ status for three different FHR-5 somamers. Relative protein levels of Complement factor H related factor 5 (FHR5) as measured by three different somamers (A-C) on the SomaScan platform from serum samples of 399 FinnGen participants recalled from four Finnish biobanks. (x-axis) reflects *CFHR5*$_{fs}$ carrier status. Univariate linear regression analyses were performed using R. In the box plots, the center line represents the median, the box limits represent the interquartile range (IQR) and the whiskers indicate the minimum and maximum values. het, heterozygotes; hom, homozygotes. RFU, relative fluorescence units.



Extended Data Figure 7. Complement activity in *CFHR5*$_{fs}$ carriers relative to AMD status. Functional analysis of CP, AP, and LP activity in serum samples from 84 recall study participants grouped into individuals with (n=40; red) and without (n=44; green) a registry-based diagnosis of AMD. (x-axis) reflects carrier status for *CFHR5* frameshift variant p.Glu163insAA. C/C, non-carriers; C/CAA, heterozygotes; CAA/CAA, homozygotes. Statistical analyses were performed with the Student's T-test or the Mann-Whitney U-test depending on whether data were normally distributed or skewed. The Kruskal-Wallis test was applied for comparisons across all three groups. Each dot represents one individual. For each group the center line represents the median, the whiskers the standard deviation. AU, arbitrary units.